Cancer Gene Therapy (2004) 11, 603–612
© 2004 Nature Publishing Group All rights reserved 0929-1903/04 $30.00
www.nature.com/cgt

Use of the EGP-2/Ep-CAM promoter for targeted expression of heterologous genes in carcinoma derived cell lines

Pamela MJ McLaughlin,[1,2] Monika Trzpis,[1,2] Bart-Jan Kroesen,[1,2] Wijnand Helfrich,[1,2] Peter Terpstra,[1,2] Wim HA Dokter,[1,2] Marcel HJ Ruiters,[1,2] Lou FMH de Leij,[1,2] and Martin C Harmsen[1,2]

[1]Department of Pathology and Laboratory Medicine, Section of Medical Biology, University Hospital Groningen, Hanzeplein 1, 9713 GZ Groningen, The Netherlands; and [2]GUIDE, Groningen University Institute for Drug Exploration, Oostersingel 59, 9713 EZ Groningen, The Netherlands.

EGP-2, also known as Ep-CAM, is expressed at high levels on the surface of most carcinomas and is therefore considered an attractive target for anticancer strategies. To explore the mechanisms regulating the expression of EGP-2, sequences 3.4 kb upstream of the transcription start site were isolated and assayed for their ability to control the expression of the EGP-2 cDNA, the green fluorescent protein, the luciferase reporter gene and the thymidine kinase and cytosine deaminase suicide genes. Expression of these chimeric constructs as assessed in a range of different cell lines was restricted to cell lines expressing EGP-2. In addition, only cells expressing EGP-2 were sensitive for gancyclovir after being transiently transfected with EGP-2 promoter-driven thymidine kinase. Deletion analyses defined 687 bp upstream as the basic proximal promoter region, which could confer epithelial-specific expression to the GFP reporter gene *in vitro*. As these EGP-2 sequences can confer promoter activity to reporter and suicide genes in an EGP-2 restricted manner, they may be useful for gene therapy of EGP-2 expressing carcinomas.
Cancer Gene Therapy (2004) **11,** 603–612. doi:10.1038/sj.cgt.7700725
Published online 9 July 2004

Keywords: transcriptional targeting; GA733-2; suicide gene therapy

Despite numerous improvements in radiological, chemotherapeutical and surgical techniques, current treatments for metastatic malignant disease are often ineffective. Therefore, new treatment strategies, which can enhance the selectivity of systemic therapy so that tumor response is increased without toxicity to normal tissue, have gained interest.[1] In this respect, gene therapy provides an attractive option, in combination with promising suicide gene/prodrug systems as effector mechanism.[2] However, a major impediment to the development of gene therapy treatments is the lack of suitable expression cassettes for directing selective transgene expression. The epithelium specific but highly abundant expression of the human epithelial glycoprotein-2 (EGP-2) makes it a useful target for carcinoma directed treatment modalities, such as EGP-2-restricted gene therapy. The 38 kb EGP-2 protein, also referred to as Ep-CAM or 17-1A, is encoded by the GA733-2 gene.[3] Although it has been described as a homotypic adhesion molecule and as ligand of the leukocyte-associated immunoglobulin-like receptor (LAIR-1) the physiological function of EGP-2 is still unclear.[4–6] Since its discovery in 1979, numerous immunotherapeutical strategies using EGP-2 as a target have been developed and are at present used in clinical settings.[7,8] Use of the EGP-2 protein's carcinoma specificity for the development of gene therapy strategies, however, has been held up by the fact that the regulatory sequences directing this specificity had not yet been characterized.

Here, we describe the isolation of the 5′ sequences from the GA733-2 gene and the identification of *cis*-acting sequences needed for selective expression of heterologous genes in EGP-2-positive cells. By deletion analysis, the basic proximal promoter region capable of directing expression in an EGP-2-restricted manner was defined. Subsequently, the EGP-2 transcriptional regulatory sequences were successfully used to direct transient, carcinoma-specific expression of the cytosine deaminase (CD) and thymidine kinase (TK) suicide genes. The use of these constructs, especially in combination with an EGP-2-specific gene delivery system as has been developed recently,[9] should enhance the safety and efficacy of vector-based carcinoma-specific gene therapy approaches.

Received April 2, 2003.
Address correspondence and reprint requests to: Dr MC Harmsen, PhD, Department of Pathology and Laboratory Medicine, Section Medical Biology, University Hospital Groningen, Hanzeplein 1, 9713 GZ Groningen, The Netherlands. E-mail: m.c.harmsen@med.rug.nl

Materials and methods

Cell culture

The GLC-1 and GLC-45 SCLC cell lines were generated at our laboratory, previously.[10] The fetal lung fibroblasts

(FLF) were isolated in 1992 under informed consent and the human umbilical veins endothelial cells (HUVEC) are isolated on a weekly bases under informed consent.[11] The human colorectal adenocarcinoma cell line SW948 (CCL 237), ovarian carcinoma cell line SKOV 3 (HTB 77), glioblastoma cell line U373 MG (HTB-17), the mouse melanoma cell line B16-F10 (CRL 6475), and the SV40 transformed simian kidney cell line COS-7 (CRL 1651) were obtained from the American Type Culture Collection (Manassas, VA). All cell lines were cultured in the recommended growth medium and maintained in a 37°C atmosphere containing 5% CO_2.

Cell transfection

The cells were transfected using the cationic lipid transfection reagent Saint (Synvolux Inc., Groningen, The Netherlands) in a six-well plate with 0.5 μg DNA per well following the manufacturer's protocol. To correct for differences in transfection efficiency, transfection experiments were carried out in triplo and repeated at least 3 times with freshly isolated plasmid DNA.[12]

Isolation and cloning of the EGP-2 promoter region

For the isolation of the EGP-2 5′ promoter region, a BAC genomic library was screened by GenomeSystem, Inc. (St Louis, MO) with a 400 bp [^{32}P] -genomic DNA fragment containing 200 bp of the 5′ region of the EGP-2 gene in addition to a part of exon 1. The probe was derived by digestion with *Sal*I/*Sac*II of the GA21726-22RS vector, kindly provided by Dr Linnenbach (Wistar Institute, Philadelphia). DNA from a positive clone was purified according to standard methods for BAC DNA isolation and analyzed by restriction mapping and Southern blot analysis. A 4.2 kb spanning *Sac*II/*Bgl*II genomic subfragment containing at least part of exon 1 and approximately 4 kb of upstream sequences was identified and isolated from the BAC vector and cloned into the *Sac*II/*Bam*HI sites of the pBluescript SK plasmid (Stratagene, Inc., San Diego, CA). This construct was then subjected to further restriction mapping and DNA sequence analysis using the thermo sequenase primer cycle sequencing kit (Amersham-Pharmacia, Biotech., Piscataway, NJ) with Cy5-labeled primers (Eurogentec, Seraing, Belgium) on the ALF-express sequencer (Amersham Pharmacia, Biotech.). DNA sequence data were managed and analyzed by the DNA Star computer program (DNA Star Inc., Madison, WI). Consensus sequences of transcription factor binding sites were identified using the TRANSFAC v3.2 database.[13] The 4.2 kb *Sac*II/*Bgl*II sequence was submitted to the gene bank accesion no. AY148099.

With *Xma*III 3.6 kb of EGP-2 promoter sequences without the ATG were seized out of the *Sac*II/*Bgl*II promoter fragment and recloned into the *Not*I site of the pBluescript SK vector (Stratagene, Inc., San Diego, CA). Via subsequent digestion with *Xho*I/*Sac*II 3.4 kb of the promoter sequence was cloned into the luciferase reporter pGL3-basic vector (Promega Inc., Madison, WI). Digestion of a pBluescript construct containing the promoter sequence in the reverse orientation with *Sac*I/*Xho*I yielded a 3.4 kb promoter fragment that was cloned into the GFP reporter plasmid pEGFP-1 (Clonetech, Palo Alto, CA).

Deletion constructs of this 3.4 kb EGP-2 promoter region containing pEGFP-N1 vector further referred to as $p39^{E}$, were generated using a *double-stranded* Nested Deletion Kit (Amersham-Pharmacia, Biotech.). Following the manufacturer's protocol deletion constructs were generated using *Bgl*II to generate the recessed 3′-ends which were filled in with thionucleotides to make them nuclease resistant and *Spe*I to create a 5′-overhanging nuclease-sensitive end. The generated EGP-2 promoter constructs chosen to be used in transfection experiments were; $p39^{E}$ (−3340/+93); $p39^{E4\text{-}7}$ (−2705/+93); $p39^{E17\text{-}1}$ (−2324/+93); $p39^{E15\text{-}2}$ (−2088/+93); $p39^{E7\text{-}2}$ (−1023/+93); $p39^{E4\text{-}1}$ (−688/+93); $p39^{E11\text{-}1}$ (−341/+93); $p39^{E12\text{-}3}$ (−57/+93), and $p39^{E12\text{-}2}$ (+9/+93). The numbers between the brackets refer to number of base-pairs relative to the putative transcription site (Fig 2). Digestion of the $p39^{E}$ vector with *Pst*I and subsequent ligation resulted in $p39^{EPstI31}$ (−177/+93) and $p39^{EPstI26}$ (−3340/−2803 and −177/+93) deletion constructs. Upon digestion with *Xca*I and *Sca*I a 306 bp fragment covering a putative ESE-1 site was deleted from the $p39^{E}$ vector resulting in $p39^{E\text{-}ESE1}$(13340/−2407 and −2097/+93). In addition, the CMV promoter was excised from the pcDNA3 vector (Invitrogen, Breda, The Netherlands) by *Bgl*II/*Hind*III digestion and cloned into the multiple cloning site of the pEGFP-1 plasmid as a positive control, yielding $p39^{CMV}$. The empty pEGFP-1 vector was used as negative control.

The EGP-2 promoter-EGP-2 cDNA construct was generated by exchanging the *Hind*III/*Xba*I luciferase gene of the pGL3-basic vector with the *Hind*III/*Xba*I EGP-2 cDNA fragment isolated from the CDM8 GA733-2 vector, kindly provided by Dr Linnenbach (Wistar Institute, Philadelphia). Subsequently, the −3967 to +74 EGP-2 promoter region was fused to the cDNA by ligation in the *Kpn*I/*Stu*I site. To clone the EGP-2 promoter upstream of the Cytosin Deaminase cDNA, the *Xba*I/*Sac*I EGP-2 promoter fragment was subcloned into the pSL301 superlinker vector (Invitrogen) and cloned in the *Spe*I/*Nhe*I sites upstream of the CD gene situated in the *Nhe*I/*Pme*I sites of the pSecTag vector (Invitrogen). The EGP-2-promoter HSV-TK construct was generated by exchanging the *Bgl*II/*Bam*HI CMV promoter fragment of the pcDNA3.1-TK construct with the *Bam*HI EGP-2 promoter fragment isolated from the $p39^{E}$ vector. Both the pSecTag-CD as well as the pcDNA3.1-TK plasmid were kindly provided by Dr H Haisma, RuG, Groningen, The Netherlands. An overview of the plasmid constructs generated is shown in Figure 1.

Analysis of expression

GFP expression was studied by microscopic and flow cytometric analysis using the Leica Quantimed 600 (Leica, Rijswijk, The Netherlands) and the Coulter Elite Cytometer (Coulter Electronics, Hilaleah, FL) and immunohistochemically using the anti-GFP polyclonal antibody (Molecular Probes, Eugene, OR). FACS results were

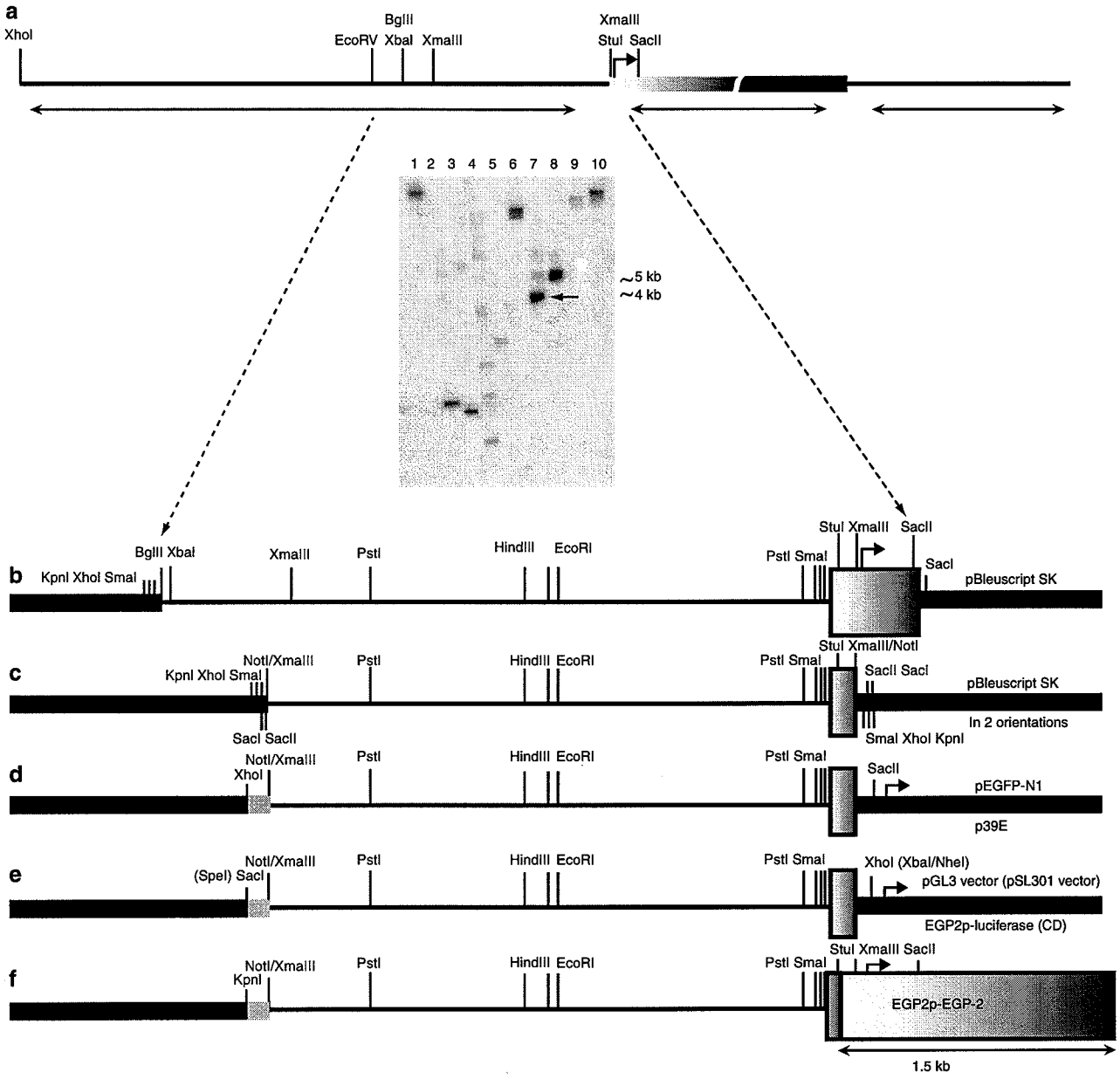

Figure 1 Identification and cloning of the EGP-2 promoter region. (**a**) Southern analysis of the BAC clone containing at least 10 kb upstream, the 14 kb EGP-2 genomic, and approximately 31 kb downstream sequences. Since a *Sac*II site is positioned approximately 40 bp downstream of the ATG, restriction analysis was carried out with *Sac*II and double digests of *Sac*II and *Hind*III, *Eco*RI, *Bam*HI, *Pst*I, *Xba*I, *Bgl*II, *Eco*RV, *Sma*I, and *Xho*I, followed by hybridization with the 5′ site of the EGP-2 genomic sequences. (**b**) Digestion of the EGP-2 promoter region with *Sac*II and *Bgl*II yielded approximately 4.2 kb of upstream sequences which were subsequently cloned in the *Sac*II and *Bgl*II sites of the pBluescript SK vector. (**c**) Upon digestion with *Xma*III, the ATG and part of the 5′ site of the upstream sequences were removed and the remaining 3.4 kb of EGP-2 promoter sequences cloned into the *Not*I site of pBluescript SK in two orientations. (**d**) Thereupon, this promoter region was cloned into the *Xho*I/*Sac*II sites of the promoterless, enhanced-GFP containing pEGFP-N1 vector yielding p39^{E}. (**e**) The same fragment but obtained from a clone containing the EGP-2 promoter in the opposite direction was cloned upon digestion by *Sac*I and *Xho*I upstream of the luciferase reporter gene in the pGL3 vector, yielding EGP2p-luciferase. Additionally, the *Sac*I/*Xho*I fragment was subcloned in the pSL301 superlinker. Digestion of this subclone with *Spe*1 and *Xba*1 and subsequent ligation of this EGP-2 promoter fragment in the *Spe*1/*Nhe*I sites of the pSeqtagCD plasmid containing the cytosine deaminase (CD) encoding cDNA, yielded EGP2p-CD. (**f**) By digestion with *Kpn*I and *Stu*I the EGP-2 promoter sequence was isolated from the pBluescript SK vector (**c**) and fused to EGP-2 cDNA leaving the first exon intact, yielding EGP2p-EGP-2.

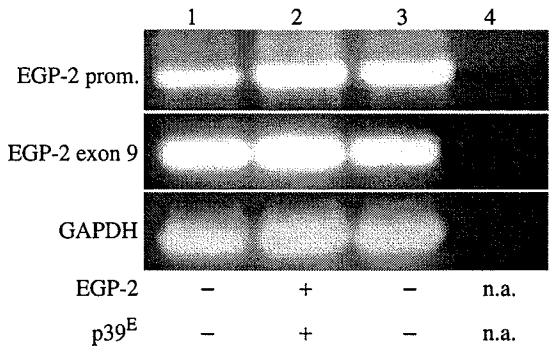

Figure 2 EGP-2 PCR analysis of the presence of the total of EGP-2 genomic sequences in the GLC-1 (1), GLC45 (2), FLF (3), cells and the H2O control (4) using primers covering the last exon, number 9, the EGP-2 promoter region, or the GAPDH gene. The presence or absence of endogenously expressed EGP-2 and expression of GFP upon transfection of the p39E vector are indicated by either a plus (+) or a minus (–) symbol.

analyzed by Winlist. 5.0 (Verity Software House, Inc., Topsham, ME) setting a gate around the background fluorescence produced by cells transfected with the promoterless pEGFP vector (R1). This setting was then used throughout the analysis of all the constructs transfected in the same cell line. The fluorescent intensity (in the FITC cannel, linMeanX) times the percentage of cells present in the R2 area when transfected with the CMV promoter inserted pEGFP ($p39^{CMV}$) was set at 100% and the relative expression level of GFP of the other constructs in the same cell line was deducted using the following equation:

$$(\text{linMeanX} \times \%\text{gate (R2)}/\text{linMeanX}\, p39^{CMV} \times \%\text{gate (R2)}p39^{CMV}) \times 100\%$$

Luciferase activity was measured using the Promega luciferase assay system (Promega Inc., Madison, WI) and light output recorded by the Anthos, Lucy 1 luminometer (Anthos Labtec Instruments, Salzburg, Austria). The relative luciferase expression (luc. exp.) was determined using the promoterless pGL3 vector as background control and the CMV promoter-driven luciferase expression ($pGL3^{CMV}$) was set at 100%. The relative expression level of the luciferase of the rest of the constructs was deducted using the following equation:

$$(\text{luc. exp.} - \text{background luc.exp.} / pGL3^{CMV}\text{luc. exp.}) \times 100\%$$

The EGP-2 expression was determined immunohistochemically using undiluted tissue-culture supernatant of the anti-EGP-2 hybridoma MOC31 (IgG1; IQ products, Groningen, The Netherlands). The antibacterial cytosine deaminase monoclonal antibody 16D8F2 kindly provided by Dr K Haack, (Chirurgische Universitätsklinik, Heidelberg, Germany), was used to identify CD-positive cells. The 9E10 c-myc peptide recognizing monoclonal antibody was used to identify TK expressing cells.[14] Horseradish-peroxidase-conjugated rabbit anti-mouse Ig and goat anti-rabbit Ig (Dakopatts, Glostrup, Denmark) were used as secondary antibodies.

EGP-2 PCR

The presence of the GA733-2 genomic sequences was determined by PCR of the last exon, exon 9 (sense strand 5′-TCAGATAAAGGAGATGGGTGAGA/antisense strand 5′-GGCAGCTTTCAATCACAAATCAG) and the promoter region from –1687 to –406 (sense strand 5′-CCGGCACTTCAACAGAATACAA/antisense strand 5′-GAACGTGGAGGCTAAAGGAAATAC). The presence of the GAPDH gene was used as control on the amount of genomic DNA (sense strand 5′-CCATCACTGCCACTCAGAAGACT/antisense strand 5′-TTACTCCTTGGAGGCCATGTAGG).

Measurement of sensitivity to gancyclovir (GCV)

SKOV-3 (EGP-2 positive), U373 MG (EGP-2 negative) and B16-F10 (EGP-2 negative) cells transiently transfected with CMV-TK, EGP-2-TK or CMV-GFP were plated 48 h after transfection in 96-wells plates at a density of 1×10^4 cells/well in recommended media. All of the experiments were performed in triplicate and repeated at least three different times. The cells were treated with various concentrations of GCV (Sigma, Bornhem, Belgium) for 4 days starting 24 h after replating. Sensitivity to GCV was evaluated using the colorimetric MTT assay as described previously.[15] A multiwell scanner was used to measure the absorbance at 570–630 nm dual wavelengths. The nontransfected untreated controls were assigned a value of 100%.

Results

Screening of a BAC genomic library (Genomic Inc.) with a 5′ GA733-2 probe yielded a positive clone, which was characterized by restriction enzyme mapping, PCR-screening, and nucleotide sequencing. To define the GA733-2 promoter region, restriction analysis with *Sac*II and double digests with *Sac*II and a number of other enzymes were carried out (Fig 1a). Since a *Sac*II site is positioned 40 bp downstream of the ATG it was shown by Southern blot hybridization with a *Sal*I/*Sac*II fragment of the 5′ GA733-2 probe that the BAC clone contained at least 10 kb of GA733-2 upstream sequences (Fig 1a, lane 1). Double digestions with *Sac*II and *Bgl*II or *Eco*RV revealed bands containing part of exon 1 and approximately 4 or 5 kb of upstream sequences, respectively (Fig 1a, lanes 7 and 8).

Sequencing of the cloned *Bgl*II/*Sac*II 5′ GA733-2 promoter region yielded 4.2 kb of sequence upstream of the longest reported 5′ untranslated region of the EGP-2 cDNA (Fig 2).[16] This upstream region lacked the canonical TATA and CAAT boxes commonly found within 100 bp upstream of the putative transcription site. However, regulatory elements competent to initiate transcription including the consensus initiator element

Table 1 Presence and absence of (epithelial-specific) consensus transcription factor recognition sequences in the proximal 5′ flanking region of the EGP-2 gene GA733-2

	Present			Absent	
Name	Consensus	Sequence	Position	Name	Consensus
Initiator (Inr)	YYANWYY	CTAGTCC	−9 (−149/−207)	TATA	TATA/TATAA
Sp-1	CCCGCC	CCCGCC	−29/−704/−1829/−1928	CCAAT	CCAAT
	GGGCGG	GGGCGG	−234	Ker1	GCCTGCAGGC
		GGCGGG	−65/−3296	E-pal	CACCTGCAGGTG
AP-1	TGASTCAG	TGACTCA	−122 (−2387)		
AP-2	GSSWGSCC	GCGTGCCC	+65		
		GGCCTCGC	+75		
Ets	A/GGAA/T	AGGAA	+248		
		TTCCT	−392/−527/−755/−2111		
		TTTCCT	−410/−1247/−1555		
		AAAGGT	−456		
		ACCTTT	−931		
		AAAGGAAG	−317		
ESE-1	CAGGAAGT	ACTTCCTG	−2263		
E-pal-like(HLH)	CANNTGCANNTG	CAACTGCAGCG	−181		
		CATCTGCACGG			

Consensus sequences identified using the TRANSFAC v3.2 database. The first two columns contain the names of the transcription factors and the consensus recognition sites present. The next two columns show the sequence of the EGP-2 promoter and the nucleotide position relative to the putative transcription start site. The following two columns, the names of transcription factors and consensus recognition sites absent. S, G or C; W, A or T; Y, C or T; N, A or C or G or T; K, G or T; M, A or C, and R, A or G. HLH, helix–loop–helix transcription factor recognition sites.

(YYANWYY) which position matches with the putative transcription initiation site and *GC* boxes were present. An E-box sequence is also present, however at position −1753. The NF-κB transcription-binding site involved in the downregulation of EGP-2 by TNFα in squamous cell carcinomas is located in proximity of both the Inr and the ATG. Screening of the proximal 5' flanking region for consensus recognition sequences of transcription factors that are known to be involved in epithelial-specific gene expression revealed putative *cis*-acting regulatory elements (Table 1). Proximal to the transcription start Ets and Sp1 transcription factor binding sites were identified. Several epithelium-specific genes, such as transglutaminase-3 (TGM-3) and keratin 18 (K18), have been shown to depend on Ets factors for epithelial cell transcription. Ets members generally cooperate with other transcription factors and it has been demonstrated that epithelial-specific expression can be directed by cooperative interactions between Sp1 and Ets transcription factors. The EGP-2 promoter also contains a consensus binding site for the activator protein-1 (AP-1) and AP-2, which have been implicated in the epithelial specific expression of the K18 and α6 integrin genes, respectively. Of further interest is the presence of an ESE-1 (epithelial-specific ets-1) consensus binding site at −2167 at the (-) strand. No consensus sequences homologous to the E-pal sequence, which has been reported to direct the epithelial-specific transcription of the E-cadherin gene, could be detected following the consensus sequence $(CANNTG)_2$. However, imperfect E-pal-like sequences harboring the CANNTG consensus binding site for the helix–loop–helix transcription factors at only one site could be found at positions −1062 and −181 of the EGP-2 promoter. No keratinocyte-specific transcription factor Ker-1 binding site often present in the promoter region of the keratin genes was detected.

To verify that the proximal 5′ flanking region contained a functional promoter capable of directing cell-specific heterologous gene expression, the reporter genes luciferase and GFP, and the EGP-2 and cytosine deaminase cDNA sequences were cloned downstream of the 3.4 kb promoter region as depicted schematically in Figure 1. The obtained plasmids were transfected into the EGP-2-positive and -negative cells and assayed for luciferase, GFP, EGP-2 or CD expression as described. Identical constructs with either the CMV promoter or without any 5′ regulatory sequences were used as positive and negative control, respectively. Table 2 shows the results of these transfection experiments. EGP-2 promoter-driven expression of luciferase, GFP and CD was observed in the EGP-2-positive carcinoma cell lines SW948 and GLC-45. EGP-2 promoter-driven expression of EGP-2 cDNA could not be determined in these cell lines due to the presence of endogenous EGP-2. No expression of luciferase, GFP, EGP-2, or CD could be observed in the EGP-2-negative GLC-1, FLF, and HUVEC cells upon transfection. In all cell lines expression of the heterologous genes could be detected when driven by the CMV promoter. Furthermore, PCR analysis on genomic DNA of the GLC1 and FLF cells, using exon 9 as well as promoter region spanning primers demonstrated that the lack of EGP-2 expression in these cells was not due to genomic deletion of the EGP-2 gene itself (Fig 3). Therefore, it is most likely that presence or absence of regulatory proteins in

Table 2 EGP-2 promoter-directed expression of the luciferase, GFP, EGP-2, and CD gene in both EGP-2-positive and -negative cells

Cell line	EGP-2 expression	EGP2p-EGP-2	EGP2p-GFP	EGP2-luciferase	EGP2p-CD
SW948	Positive	n.a.	+	+	+
GLC-45	Positive	n.a.	+	+	+
GLC-1	Negative	–	–	–	–
FLF	Negative	–	–	–	–
HUVEC	Negative	–	–	–	–
COS-7	n.a.	+	+	+	+

EGP-2, GFP, luciferase, and CD expression directed by the EGP-2 5′ proximal promoter region in the EGP-2 expressing carcinoma cell lines, SW948 and GLC-45, and in the EGP-2 negative cell lines GLC-1, fetal lung fibroblast (FLF), and human umbilical cord-derived endothelial cells (HUVEC). COS-7 cells were used to establish the functionality of the generated constructs. n.a. not applicable. Table represents the results of three independent transfection experiments with freshly isolated DNA.

these cells causes the lack of EGP-2 expression. These results show that the 3.4 kb EGP-2 5′ regulatory sequences contain a functional promoter and suggest that the activity of this promoter is restricted to cells expressing EGP-2 due to cell type-specific gene transcription.

To identify regions in the EGP-2 promoter responsible for the observed epithelial cell-specific expression, deletion analyses were performed. Deletion constructs of the $p39^{E}$ GFP reporter plasmid containing −3340 nucleotides upstream of the transcription initiation site, the transcription initiation site, and +93 nucleotides of the untranslated region of exon 1 (Fig 1d) were generated and transfected into COS-7, SW948, FLF, and HUVEC cells and analyzed for GFP expression. All constructs generated were sequenced and all of the sequences were identical to the cloned and sequenced 5′ region of EGP-2 genomic DNA as depicted in Figure 2. As shown in Figure 4, minimal promoter activity was confined to −177 bp upstream sequences, since transfection of construct $p39^{EPstI}$ showed maximal expression of GFP in both the COS-7 and SW948 cell lines. This construct contains the initiation of transcription consensus, two putative Sp-1 and AP-2 binding sites, an AP-1 and NF-κB consensus sequence, and contains half of the first putative E-pal sequence. Deletion of the AP-1 and one of the Sp-1 bindings sites reduced the promoter activity to approximately 60% of the maximum EGP-2 promoter activity ($p39^{E12\text{-}3}$), whereas removal of the Inr further reduced the promoter activity with approximately 25% ($p39^{E12\text{-}2}$). GFP expression directed by −351 ($p39^{E11\text{-}1}$) promoter sequences was not epithelial specific since promoter activity of constructs covering this promoter region was also observed in the FLF and HUVEC. Transfection of an additional −337 bps ($p39^{E4\text{-}1}$) starting at −688, did confine epithelial-specific expression to the GFP reporter construct since only limited GFP expression could be detected in FLF and HUVEC upon transfection, whereas maximal expression of GFP was observed in the COS-7 and SW948 cell lines. Deletion of -2803 to -177 of the EGP-2 promoter region resulted in maximal GFP expression in all cell lines tested implicating that this region harbors epithelial-specific transcriptional elements. Complete silencing of GFP expression in FLF cells was achieved using −2324 bp of upstream sequences ($p39^{E17\text{-}1}$). The construct $p39^{E17\text{-}1}$ contains additional 246 bp of EGP-2 promoter sequences that covers a consensus-binding site for an epithelial specific ets-1 (ESE-1). Removal of this ESE-1 site and the nearby AP-1 consensus binding site by *Xcal* and *Scal* digestion demonstrated that this part of the 3.4 kb EGP-2 promoter sequences was important for maintenance of the cell type specificity in FLF cells. However, since expression was not completely restored in the FLF and not regained by this construct in HUVEC other specific regulation factors, as yet unknown, must be of importance as well.

To evaluate the EGP-2 promoter for its capacity to induce tumor cell-specific death, we cloned the 3.4 kb EGP-2 promoter sequences upstream of the HSV-TK gene. As a positive control the CMV-promoter-driven HSV-TK construct was used whereas the CMV promoter-driven GFP construct functioned as negative control and as indicator of the number of cells transfected. Cells were transiently transfected as this mimics best actual EGP-2 promoter-driven gene therapy of EGP-2 expressing carcinomas in humans. GCV sensitivity induced by expression of the CMV-TK construct and determined by MTT could not be measured in cell lines with a transfection efficiency beneath approximately 30%. Maximal CMV-TK induced GCV sensitivity was observed in SKOV-3, U373 MG and B16-F10 cell lines at a concentration of 0.1 mM or higher and appeared to correlate with the transfection efficiency (Fig 5). Expression of TK in these cell lines was immunohistochemically confirmed (data not shown). Barely any sensitivity to GCV was observed in these cell lines upon transfection with the CMV-GFP construct. Upon transfection with the EGP-2-TK construct sensitivity to GCV could only be observed in the EGP-2 expressing ovarian carcinoma cell line SKOV-3. This demonstrates that the 3.4 kb EGP-2 promoter is sufficient enough to drive EGP-2 expressing carcinoma-cell-specific killing via the HSV-TK suicide gene/GCV prodrug system.

Discussion

Here, we report the isolation and characterization of 3.4 kb of *cis*-acting DNA sequences that control the expression of the gene encoding the pancarcinoma

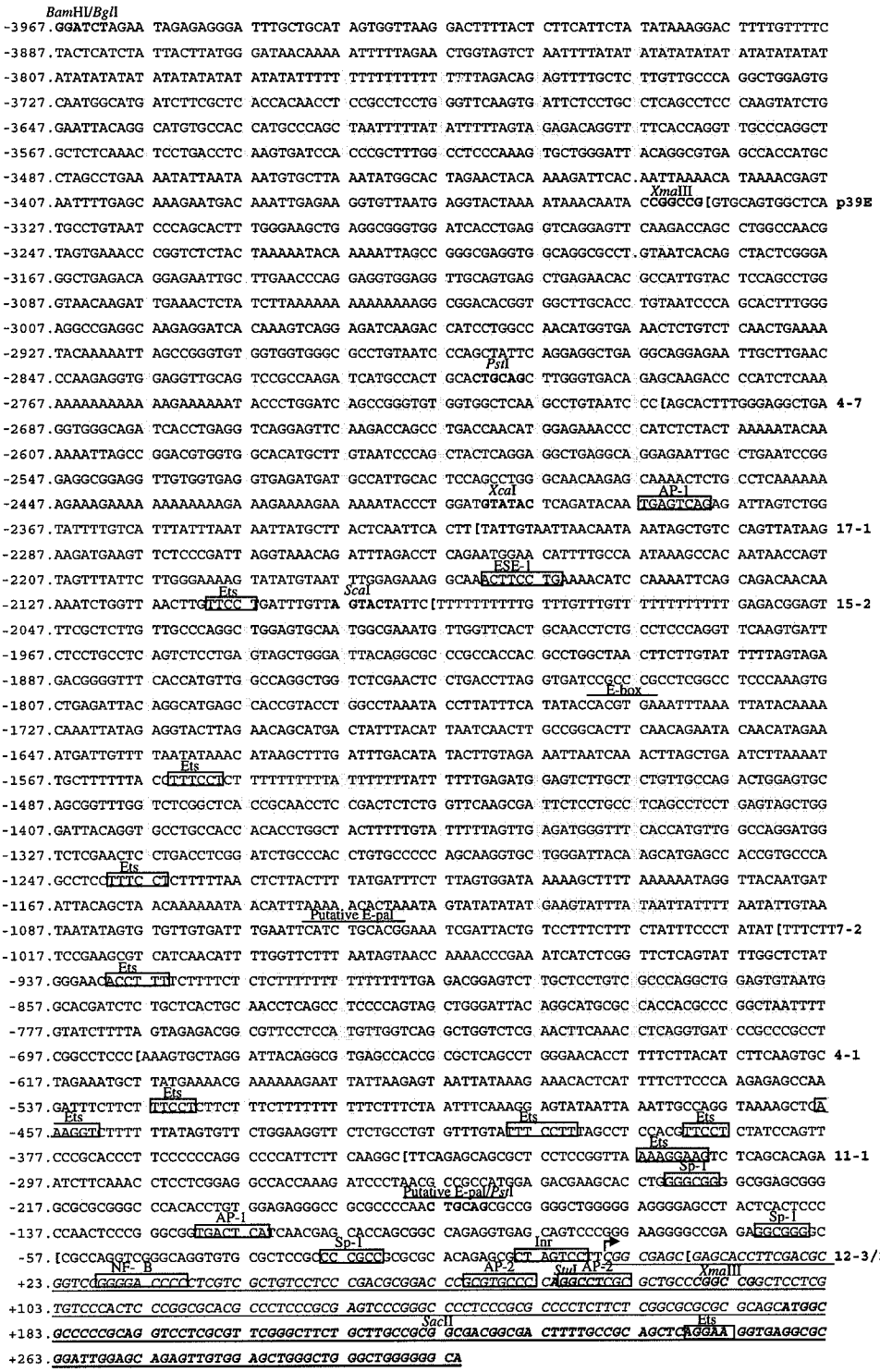

Figure 3 Organization of the EGP-2 5′ regulatory region plus exon 1. Putative binding sites for Sp1, AP1, AP2, Ets and ESE-1 transcription factors as well as the initiator element are boxed. The bend arrow depicts the putative transcription start corresponding to the 5′ end of the cDNA clone with the longest reported 5′-untranslated region[14] is underlined. The ATG start codon and protein-encoding region is depicted bold, underlined in italic. Gray shadowed sequences were defined as repeat sequences using the mask repeat sequence program. The brackets depict the start of the deletion construct named at the side. Nucleotide positions are numbered to the right with respect to the transcription initiation site at +1.

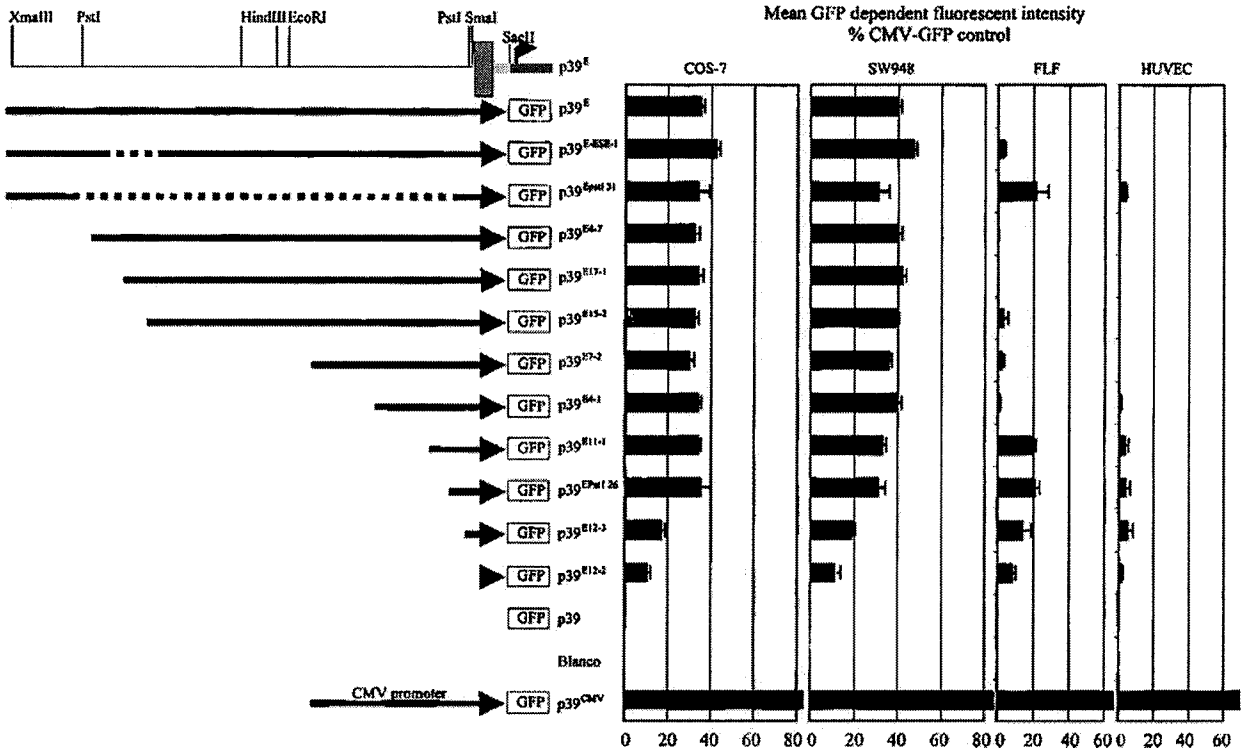

Figure 4 Diagram of GFP expression of the EGP-2 promoter-deletion constructs in the EGP-2-positive cell line SW948 and the EGP-2-negative cell lines FLF and HUVEC. COS-7 cells were used to test the functionality of the constructs generated. The top line represents the EGP-2 5′ sequences. The arrow indicates the orientation of the sequences and the dashed lines represent internal deletions. Flow cytometric measurements of the mean fluorescent intensity of GFP expression driven by the CMV promoter ($p39^{CMV}$) was set at 100% and the GFP expression of the promoterless GFP vector (p39) was set as background. Percent $p39^{CMV}$ GFP values are shown as the average ±SE of the mean. Error bars (SEM) are present on all bar graphs except $p39^{CMV}$.

associated-antigen EGP-2, also known as GA733-2, Ep-CAM or 17-1A, and the successful use of this promoter region to direct epithelial-specific expression of a number of heterologous genes.

Analysis of the EGP-2 5′ regulatory sequences, revealed several homologies to known transcriptional regulatory sequences. Although no TATA or CAAT box could be distinguished, other typical eukaryotic promoter elements like an initiator (Inr) consensus and GC boxes are present.[17] Previous studies showed that EGP-2 is down-regulated upon treatment with TNFα, TPA and IFNα, via NF-κB.[18,19] However, we screened the proximal 5′ flanking region especially for consensus recognition sequences of transcription factors known to be involved in epithelial-specific gene expression. The GC boxes upstream of the Inr confirm to a consensus-binding site for the transcription factor Sp1. Sp1 can initiate transcription by recruiting the RNA polymerase holoenzyme to the promoter and has been described to be involved in epithelial-specific expression upon interaction with an Ets consensus-binding element.[20,21] Although numerous Ets consensus-binding sites are present throughout the EGP-2-promoter sequences no Ets consensus-binding site is present near the two Sp1 sites closest to the Inr. Sp1 and Ets consensus-binding sites located in proximity can be found more upstream. Examination of the EGP-2 promoter for the presence of other known epithelial specificity-conveying consensus sequences yielded two imperfect E-pal-like sequences and an ESE-1 consensus-binding site. The E-pal sequence contributes to the epithelial-specific expression of the E-cadherin gene and E-pal-like sequences are involved in the epithelial-specific expression of MUC1 and Integrin α6.[22–24] However, these sequences all harbor two CANNTG consensus-binding sites for helix–loop–helix transcription factors and not one, like the two putative E-pal-like sequences found in the EGP-2 promoter. However, only one-half of the E-pal sequence is necessary for its function.[25]

The EGP-2 regulatory sequences analyzed here are capable of directing heterologous gene expression in epithelial cells normally expressing EGP-2 but not in non-EGP-2-expressing cells. By deletion analysis it was established that 177 bp of the 5′ flanking sequence are sufficient to give maximal promoter activity, whereas 687 bp of the 5′ flanking region are sufficient to confer epithelial specificity. In the FLF cells, expression directed by the EGP-2 promoter seems to depend on a combination of yet undefined cell type-specific elements and the ESE-1 element present. The ESE-1 binding site, situated

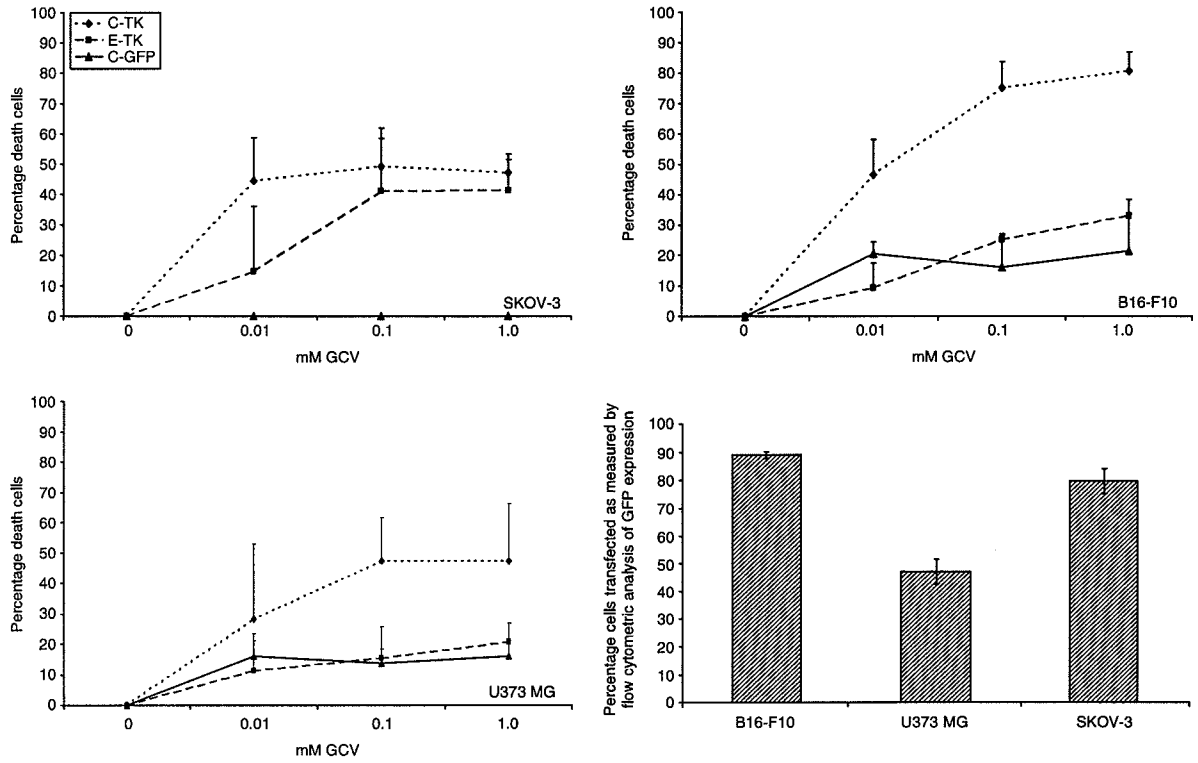

Figure 5 EGP-2 promoter-driven TK expression sensitizes specifically an EGP-2 expressing cell line to GCV. SKOV-3 (EGP-2 positive), B16-F10 and U373 MG (both EGP-2 negative) cells transfected with CMV-TK, EGP-2 TK or CMV-GFP constructs were cultured in medium containing 0, 0.01, 0.1 or 1.0 mM GCV for 4 days after which cell death was determined. Presented are the percentages of dead cells compared with untreated cells, the mean of triplicate samples. Representative results shown are from one of at least three separate experiments. All cells transfected with CMV-TK could be sensitized to GCV above a concentration of 0.1 mM, although the amount of killing appeared to depend on the number of cells transfected per cell line. Cells transfected with CMV-GFP showed some background level of sensitivity to GCV. Only in SKOV-3 cells significant GCV sensitivity was observed upon transfection with the EGP-2-TK construct.

2167 bp upstream of the Inr, is of particular interest since ESE1 is expressed in a variety of simple and stratified epithelia with a high expression in the epithelial lining of the gastrointestinal tract.[26] This distribution pattern strongly resembles the EGP-2 distribution pattern found in the EGP-2 expressing transgenic mouse model, we generated using the total of 55 kb EGP-2 encoding and regulatory genomic sequences from the isolated BAC clone.[27]

The TK gene is capable of converting the nontoxic prodrug GCV into the toxic monophosphate form leading to the death of TK expressing cells. We show that the EGP-2 promoter sequences are capable of directing the TK gene expression in an EGP-2 expressing-carcinoma-specific manner. However, no bystander effect, which is often reported as an important aspect of the HSV-TK/GCV antitumor therapy, could be observed in the cell lines used.[28] In applying enzyme/prodrug therapy it is crucial to specifically damage tumor cells without affecting normal cells. The EGP-2 regulatory sequences described here are sufficient to enhance the specificity in carcinoma-directed suicide gene therapy. Use of these sequences should enhance the safety and efficacy of vector-based carcinoma-specific gene therapy approaches.

References

1. Haviv YS, Curiel, DT. Conditional gene targeting for cancer gene therapy. *Adv Drug Deliv Rev.* 2001;53:135–154.
2. Denny WA. Prodrug strategies in cancer therapy. *Eur J Med Chem.* 2001;36:577–595.
3. Balzar M, Winter MJ, de Boer CJ, Litvinov SV. The biology of the 17-1A antigen (Ep-CAM). *J Mol Med.* 1999;77:699–712.
4. Litvinov SV, Velders MP, Bakker HA, et al. Ep-CAM: a human epithelial antigen is a homophilic cell–cell adhesion molecule. *J Cell Biol.* 1994;125:437–446.
5. Meyaard L, van der Vuurst de Vries AR, de Ruiter T, et al. The epithelial cellular adhesion molecule (Ep-CAM) is a ligand for the leukocyte-associated immunoglobulin-like receptor (LAIR). *J Exp Med.* 2001;194:107–112.
6. Linnenbach AJ, Wojcierowski J, Wu SA, et al. Sequence investigation of the major gastrointestinal tumor-associated antigen gene family, GA733. *Proc Natl Acad Sci USA.* 1989;86:27–31.
7. Herlyn M, Steplewski Z, Herlyn D, Koprowski H. Colorectal carcinoma-specific antigen: detection by means of monoclonal antibodies. *Proc Natl Acad Sci USA.* 1979;76:1438–1452.
8. Riethmuller G, Holz E, Schlimok G, et al. Monoclonal antibody therapy for resected Dukes' C colorectal cancer:

seven-year outcome of a multicenter randomized trial. *J Clin Oncol.* 1998;16:1788–1794.
9. Heideman DA, Snijders PJ, Craanen ME, et al. Selective gene delivery toward gastric and esophageal adenocarcinoma cells via EpCAM-targeted adenoviral vectors. *Cancer Gene Ther.* 2001;8:342–351.
10. de Leij LF, Postmus PE, Buys CH, et al. Characterization of three new variant type cell lines derived from small cell carcinoma of the lung. *Cancer Res.* 1985;12:6024–6033.
11. Mulder AB, Blom NR, Smit JW, et al. Basal tissue factor expression in endothelial cell cultures is caused by contaminating smooth muscle cells. Reduction by using chymotrypsin instead of collagenase. *Thromb Res.* 1995;80: 399–411.
12. Farr A, Roman A. A pitfall of using a second plasmid to determine transfection efficiency. *Nucleic Acids Res.* 1992;20:920.
13. Werner T. Computer-assisted analysis of transcription control regions. Matinspector and other programs. *Methods Mol Biol.* 2000;132:337–349.
14. Evan GI, Lewis GK, Ramsey G, Bishop JM. Isolation of monoclonal antibodies specific for human c-myc proto-oncogene product. *Mol Cell Biol.* 1985;5:3610–3616.
15. Mosmann T. Rapid colorimetric assay for cellular growth and survival: application to proliferation and cytotoxicity assays. *J Immunol Methods.* 1983;16:55–63.
16. Szala S, Froehlich M, Scollon M, et al. Molecular cloning of cDNA for the carcinoma-associated antigen GA733-2. *Proc Natl Acad Sci USA.* 1990;87:3542–3546.
17. Struhl K. Fundamentally different logic of gene regulation in eukaryotes and prokaryotes. *Cell.* 1999;98:1–4.
18. Gires O, Kieu C, Fix P, et al. Tumor necrosis factor α negatively regulates the expression of the carcinoma-associated antigen epithelial cell adhesion molecule. *Cancer.* 2001;92:620–628.
19. Gires O, Eskofier S, Lang S, et al. Cloning and characterisation of a 1.1 kb fragment of the carcinoma-associated epithelial cell adhesion molecule promoter. *Anticancer Res.* 2003;23:3255–3262.
20. Smale ST, Schmidt MC, Berk AJ, Baltimore D. Transcriptional activation by Sp1 as directed through TATA or initiator: specific requirement for mammalian transcription factor IID. *Proc Natl Acad Sci USA.* 1990;87:4509–4513.
21. Lee JH, Jang SI, Yang JM, et al. The proximal promoter of the human transglutaminase 3 gene. Stratified squamous epithelial-specific expression in cultured cells is mediated by binding of Sp1 and ets transcription factors to a proximal promoter element. *J Biol Chem.* 1996;271:4561–4568.
22. Behrens J, Lowrick O, Klein-Hitpass L, Birchmeier W. The E-cadherin promoter: functional analysis of a G–C-rich region and an epithelial cell-specific palindrome regulatory element. *Proc Natl Acad Sci USA.* 1991;88:11495–11499.
23. Kovarik A, Peat N, Wilson D, et al. Analysis of the tissue-specific promoter of the MUC1 gene. *J Biol Chem.* 1993;268:9917–9926.
24. Lin CS, Chen Y, Huynh T, Kramer R. Identification of the human alpha6 integrin gene promoter. *DNA Cell Biol.* 1997;16:929–937.
25. Hennig G, Lowrick O, Birchmeier W, Behrens J. Mechanisms identified in the transcriptional control of epithelial gene expression. *J Biol Chem.* 1996;271:595–602.
26. Oettgen P, Alani RM, Barcinski MA, et al. Isolation and characterization of a novel epithelium-specific transcription factor, ESE-1, a member of the ets family. *Mol Cell Biol.* 1997;17:4419–4433.
27. McLaughlin PM, Harmsen MC, Dokter WH, et al. The epithelial glycoprotein 2 (EGP-2) promoter-driven epithelial-specific expression of EGP-2 in transgenic mice: a new model to study carcinoma-directed immunotherapy. *Cancer Res.* 2001;61:4105–4111.
28. Kuriyama S, Nakatani T, Masui K, et al. Bystander effect caused by suicide gene expression indicates the feasibility of gene therapy for hepatocellular carcinoma. *Hepatology.* 1995;22:1838–1846.